World Cup Qatar 2022 predictions: semifinals

Leonardo Egidi, Vasilis Palaskas - Mail: legidi@units.it, vasilis.palaskas94@gmail.com

12 December 2022

Contents

The statistical model (in brief) **1**

Semifinals predictions (13-14 December) **2**

World cup winning probabilities **3**

The statistical model (in brief)

We use a **diagonal-inflated Bivariate-Poisson model with dynamic team-specific abilities** for the attack and the defence. Let (X_i, Y_i) denote the random number of goals scored by the home and the away team in the i-th game, $i = 1, \ldots, n$, respectively. ranking denotes the Coca-Cola FIFA ranking at October 6th, 2022, whereas att and def denote the attack and the defence abilities, respectively.

$$
\begin{aligned}
(X_i, Y_i) \ &\sim \ \begin{cases} (1-p)\text{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) & \text{if } x \neq y \\ (1-p)\text{BP}(x_i, y_i|\lambda_1, \lambda_2, \lambda_3) + pD(x, \eta) & \text{if } x = y, \end{cases}, && (1)\\
\log(\lambda_{1i}) &= \ \text{att}_{h_i,t} + \text{def}_{a_i,t} + \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}) && (2)\\
\log(\lambda_{2i}) &= \ \text{att}_{a_i,t} + \text{def}_{h_i,t} - \frac{\gamma}{2}(\mathsf{ranking}_{h_i} - \mathsf{ranking}_{a_i}), \quad i = 1, \ldots, n \text{ (matches)}, && (3)\\
\log(\lambda_{3i}) &= \ \rho, && (4)\\
\text{att}_{k,t} &\sim \ \mathcal{N}(\text{att}_{k,t-1}, \sigma^2), && (5)\\
\text{def}_{k,t} &\sim \ \mathcal{N}(\text{def}_{k,t-1}, \sigma^2), && (6)\\
\rho, \ \gamma &\sim \mathcal{N}(0, 1) && (7)\\
p &\sim \text{Uniform}(0, 1) && (8)\\
\sum_{k=1}^{n_t} \text{att}_{k,} = 0,& \ \sum_{k=1}^{n_t} \text{def}_{k,} = 0, \quad k = 1, \ldots n_t \text{ (teams)}, \ t = 1, \ldots, T \text{ (times)}. && (9)
\end{aligned}
$$

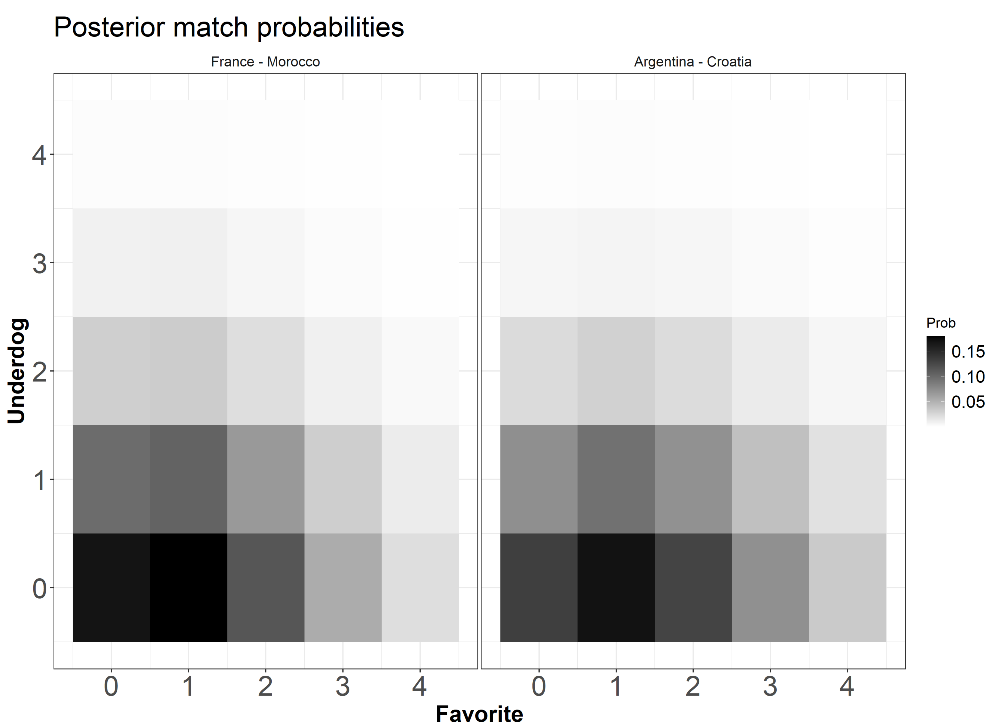
Lines (1) displays the likelihood's equations (diagonal inflated bivariate Poisson); lines (2)-(4) display the log-linear models for the scoring rates λ_1, λ_2 and the covariance parameter λ_3; lines (5)-(6) display the dynamic prior distributions for the attack and the defence parameters, respectively; lines (7)-(8) display prior distributions for the other model parameters; line (9) displays the sum-to-zero identifiability constraints. Model fitting has been obtained through the Hamiltonian Monte Carlo sampling, 2000 iterations, 4 chains using the `footBayes` `R` package (with the underlying `rstan` package). The historical data used to fit the models come from *all the international matches played during the years' range 2018-2022 additionally to the group-stage, the round of 16 and the quarter of finals matches of World Cup 2022.*

The idea is to provide a dynamic predictive scenario: at the end of each match-day, the model will be refitted to predict the remaining matches. Concerning the prediction of matches for the semifinals of WC 2022, our dynamic priors for both the attacking and defensive net abilities of the competing teams are focused on their previous five matches (group stage, round of 16, and quarter of finals matches) in such a way the three first matches contribute as a unique temporal period, and not as three distinct periods, whereas round of 16 and quarter of finals contribute as a second temporal period.

Semifinals predictions (13-14 December)

Posterior matches probabilities from the posterior predictive distribution of the model above are displayed in the table below. **mlo** denotes the most likely exact outcome (in parenthesis, the corresponding posterior probability). Darker regions in the plots below denote more likely outcomes: on the x-axis the favorite team goals, on the y-axis the underdog team goals.

favorite	underdog	favorite win	draw	underdog win	mlo
Argentina	Croatia	0.580	0.263	0.157	1-0 (0.17)
France	Morocco	0.503	0.302	0.195	1-0 (0.18)



World cup winning probabilities

We report the World Cup 2022 winning probabilities for the four teams involved in the two semifinals obtained by running some ahead simulations:

Team	Win prob.
Argentina	45%
France	35%
Croatia	12%
Morocco	8%